

## STAT100 Problem Set 8: Hypothesis Testing Involving One Sample

You need to submit a Word document or PDF for this assignment. Make sure you do the following:

1. Upload only one file/document in ELMS for Problem Set 8
2. Include your name in the document in the upper left-hand corner. Under your name, write STAT 100 and your section number. Write Problem Set 8 centered on the page.
3. Number and letter your answers to the questions accordingly.
4. Carefully read all problems and follow all instructions.
5. Upload the assignment in ELMS before the deadline of Sunday 5/1 at 11:59 PM otherwise it is considered late. Make sure you save your document on your computer or email it to yourself so that you keep an electronic copy.

Students should refer to the Tutorial for Problem Set 8 as they are working on this problem set. All Tutorials for Problem Sets can be found in the STAT100 ELMS course under Modules.

For this assignment you need to use the *BodyFatPercentage.RData* file, which includes data collected for a study on obesity in adolescent girls. Below is a brief summary of the data set:

*A medical researcher is studying obesity in adolescent girls. The researcher collected body mass index (BMI), body fat percentage, and other personal variables from 92 adolescent girls. The data collected by the researcher is a random sample representative of the population of all adolescent girls. The following variables were collected:*

- BMI: the body mass index of each girl
- Pct\_Fat: the body fat percentage of each girl
- Activity: the activity level for each girl (high, medium, or low)
- Menarche: whether the girl has achieved menarche (Yes or No)

***We do not know the population standard deviation for adolescent girls' BMI or body fat percentage.***

Problem Set 8 has two questions worth 25 points. Read each question carefully and follow all instructions. Please follow these instructions for providing your responses:

- For #1.a., #1.g., #2.a. and #2.e., you need to provide the R code that you use to generate output and provide the output displayed in the RStudio Console. In addition, you MUST include your name in the comments for the R code for #1.a., #1.g., #2.a. and #2.e.
- For #1.d., #1.e., and #2.d., you can provide R code and RStudio output, but it is not required.
- For all other questions on Problem Set 8, you should type your responses directly in the document you submit for Problem Set 8. You should NOT provide R code or RStudio output for those questions. If requested, be sure to show all work to receive full credit.

1. (13 points) You will use sample data for the activity level (*Activity*) variable from the *BodyFatPercentage* data file to perform a hypothesis test.
  - a. Using RStudio, generate two tables for the *Activity* variable: one table showing frequencies and one table showing percentages. **This is exactly the same question asked in Problem Set 7, #2.a.; you can use the same image for this question that you used for Problem Set 7, #2.a.** *Refer to the Tutorial for Problem Set 8 for guidance on generating the tables.* In the document you upload for this assignment, for 1.a. include an image showing the R code you used, including comments, and with the output from the RStudio Console.  
**IMPORTANT INSTRUCTIONS: To receive credit, you MUST include your name in the comments for the R code for #1.a.**
  - b. We want to use the *BodyFatPercentage* data set to test the claim that more than 77.5% of all adolescent girls have a have a “medium” activity level. The null hypothesis is  $H_0 : p = .775$ . State the alternate hypothesis for a one-sided test that would test if more than 77.5% of all adolescent girls have a have a “medium” activity level. **For 1.b., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8.** *Be sure to use the correct notation and symbols.*
  - c. Check that the conditions are met for this hypothesis test. You can assume that the *BodyFatPercentage* data set is a random sample. **For 1.c., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8.** *Show all work and be sure to use the correct notation and symbols.*
  - d. Calculate the test statistic for this hypothesis test and show all work. **For 1.d., you can use RStudio but cannot use the *prop.test* command.** *Refer to the Tutorial for Problem Set 8 for guidance on calculating the test statistic in RStudio without using the *prop.test* command.*
  - e. Determine the p-value for this test. **For 1.e., you can use RStudio but cannot the *prop.test* command.** *Refer to the Tutorial for Problem Set 8 for guidance on calculating the p-value in RStudio without using the *prop.test* command.* If you don't use RStudio for 1.e., you can calculate the p-value using the STAT100 Z Table and state the p-value as a decimal rounded to four decimal places.

- f. Based on your p-value from 1.e., draw a conclusion for this hypothesis test and give the reason for the conclusion. Use a level of significance of  $\alpha = .05$ . **For 1.f., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8. Be sure to give the reason for your conclusion.**
- g. Use the *prop.test* command in RStudio to perform this hypothesis test. *Refer to the Tutorial for Problem Set 8 for guidance on using the prop.test command in RStudio.* In the document you upload for this assignment, for 1.g. include an image showing the R code you used to perform the hypothesis test, including comments, and with the output from the RStudio Console. **IMPORTANT INSTRUCTIONS: To receive credit, you MUST include your name in the comments for the R code for #1.g.**
2. (12 points) You need to use information about the body mass index (*BMI*) variable from the *BodyFatPercentage* data file to perform a hypothesis test.
- a. Using RStudio, generate output showing the sample mean, sample standard deviation, and sample size for the *BMI* variable. **Note that this is NOT the same question asked in Problem Set 7, #1.a. (on Problem Set 7, #1.a., students generated output for the *Pct\_Fat* variable).** *Refer to the Tutorial for Problem Set 8 for guidance on generating the appropriate output.* In the document you upload for this assignment, for 2.a. include an image showing the R code you used to generate the sample statistics, including comments, and with the output from the RStudio Console. **IMPORTANT INSTRUCTIONS: To receive credit, you MUST include your name in the comments for the R code for #2.a.**
- b. We want to use the *BodyFatPercentage* data to test the claim that the mean BMI for all adolescent girls is less than 21.5. The null hypothesis is  $H_0 : \mu = 21.5$ . State the alternate hypothesis for a one-sided test that would test if the mean BMI for all adolescent girls is less than 21.5. **For 2.b., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8. Be sure to use the correct notation and symbols.**
- c. Check that the conditions are met for this hypothesis test. You can assume that the *BodyFatPercentage* data set is a random sample. **For 2.c., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8. Show all work and be sure to use the correct notation and symbols.**
- d. Calculate the test statistic for this hypothesis test and show all work. **For 2.d., you can use RStudio but cannot the *t.test* command.** *Refer to the Tutorial for Problem Set 8 for guidance on calculating the test statistic in RStudio without using the *t.test* command.*

- e. Use the `t.test` command in RStudio to perform this hypothesis test. *Refer to the Tutorial for Problem Set 8 for guidance on using the `t.test` command in RStudio.* In the document you upload for this assignment, for 2.e. include an image showing the R code you used to perform the hypothesis test, comments for the R code, and the output from the RStudio Console. **IMPORTANT INSTRUCTIONS: To receive credit, you MUST include your name in the comments for the R code for #2.e.**
- f. Based on the output you provide for 2.e., identify the p-value for this test. *State the p-value as a decimal rounded to four decimal places.*
- g. Based on your p-value from 2.f., draw a conclusion for this hypothesis test and give the reason for the conclusion. Use a level of significance of  $\alpha = .05$ . **For 2.g., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8. Be sure to give the reason for your conclusion.**
- h. **If this were a two-sided test** to test if the if the mean BMI for all adolescent girls is **not equal to 21.5**, and still using the same sample data (same values of sample mean, sample standard deviation and sample size), what would be the p-value of the two-sided test? **For 2.h., do NOT provide R code or RStudio output; you should type your response directly in the document you submit for Problem Set 8. State the p-value as a decimal rounded to four decimal places and explain how you determined the p-value (explain your reasoning).**